A Linked Data Approach to Sentiment and Emotion Analysis of Twitter in the Financial Domain

J. Fernando Sánchez-Rada[1], Marcos Torres[1], Carlos A. Iglesias[1], Roberto Maestre[2], and Esther Peinado[2]

[1] Universidad Politécnica de Madrid,
ETSI Telecomunicación, Avda. Complutense, 30, 28040 Madrid, Spain
[2] Paradigma Labs, Paradigma Tecnológico,
Avda. Europa, 26, Pozuelo de Alarcón, 28224 Madrid, Spain

Abstract. Sentiment analysis has recently gained popularity in the financial domain thanks to its capability to predict the stock market based on the wisdom of the crowds. Nevertheless, current sentiment indicators are still silos that cannot be combined to get better insight about the mood of different communities. In this article we propose a Linked Data approach for modelling sentiment and emotions about financial entities. We aim at integrating sentiment information from different communities or providers, and complements existing initiatives such as FIBO. The approach has been validated in the semantic annotation of tweets of several stocks in the Spanish stock market, including its sentiment information.

Keywords: linked data, semantic, finance, sentiment analysis, emotions

1 Introduction

The proliferation of user generated content in web sites and social networks, such as Facebook, TripAdvisor or Twitter, has lead to an increased awareness of the power of social networks for expressing opinions about products, services and even disasters. These so-called social sensors enable real time indexing of the social web with the aim of providing insight about the structure and activity of social networks. They provide a vast array of application possibilities, from monitoring brands or products to become early disaster warning systems [1].

In the financial field, social sensors can provide additional valuable information that complements other sources of information used in fundamental analysis, such as financial newspapers. In particular, sentiment analysis has been one of the most popular technologies to measure the investment mood. The sentiment stock market indicator has become a popular indicator that is provided together with the classical fundamental and technical stock market indicators [2]. Several websites provide the investor emotion index[3] or their sentiment, like AII Investor

[3] Market Emotion by CNN Money available at `http://money.cnn.com/data/fear-and-greed/`

Sentiment Survey[4], StockMarketSensor[5], or SentimentTrader[6], just to name a few.

In addition, recent research has shown that sentiment expressed in microblogging sites such as Twitter can be applied to predict daily changes in stock values [3,4].

Linked Data is another valuable resource that can provide financial analysts with an integration of available data sources in their activity [5]. Linked Data can provide a wide array of opportunities in the financial field. As reported by O'Riain et al. [6], depending on the information consumer needs, the integration and augmentation of financial information can lead to a significant benefit for financial and business analysis in tasks such as competitive analysis, fraud detection or figures comparison. It is also worth mentioning the recent trend towards open government and eGovernment data initiatives for public sector information, statistics data and economic indicators. The current status is promising, with a large volume of financial and economic data sets already available. Several researchers have shown this potential for different use cases, such as cross-lingual query of financial and business data from multiple sources [7,8], using social media in investment decisions [9,10] or enriching corporate financial reporting [11].

The aim of this article is the application of a Linked Data approach to expressing sentiments and emotions about financial concepts, which financial analysts can use to combine opinions expressed in different social media sites.

The article is arranged as follows. Sect. 2 gives an overview of the vocabularies we have defined for modelling sentiment and opinions as well as its interlinking with financial vocabularies such as FIBO [12]. Sect. 3 outlines our system design. Sect. 4 provides an overview of our experimental design and results. Sect. 5 expresses our conclusions and a brief discussion of future directions for this line of research.

2 Modeling Sentiment and Emotions as Linked Data

This section provides insight about the potential of Linked Data for accessing, interlinking and reasoning about business data sources. To leverage that power, it is necessary to have a robust representation model for sentiment in the financial context. Rather than creating an ad-hoc model, the Linked Data approach is to look for models for each domain and connect them. In particular, we will need a model for financial entities, a model for sentiment analysis results, and a model for microblogging messages. The following sections review the models (also referred to as ontologies or vocabularies) available in these domains, and Sect. 2.3 exemplifies the use of the final integrated model.

[4] AII Investor Sentiment Survey available at `http://www.aaii.com/sentimentsurvey`
[5] Available at `http://www.stockmarketsensor.com/`
[6] Available at `http://www.sentimentrader.com/`

2.1 Linked Data in the Financial Domain

Financial Industry Business Ontology (FIBO) [12] is a collaborative industry initiative to describe financial data standards using semantic technology. FIBO has been authored by Enterprise Data Management (EDM) council under the technical governance of the Object Management Group (OMG). FIBO has two distinct aspects: a business ontology and a presentation for business readability. FIBO is released in discrete ontologies by subject area: (i) Business Entities; (ii) Security, Loans, Derivatives and (iii) Corporate Actions and Transactions. At the time of this writing, only the first specification for Business Entities has been made public. The specification identifies a taxonomy of basic entities: Human Being, Legal Person, Organization and Legal Entity. This taxonomy is extended with other derived entities, such as Minor, Natural Person, Artificial Person (Company Limited by Guarantee, Legally Incorporated Partnership, Foundation or Incorporated Company), Formal Organization (Trust, Partnership or Incorporated Company) and Informal Organization. In addition, the ontology models concepts such as control and ownership.

Financial Exchange Framework Ontology (FEF) [13] is an ontology defined by International Financial Information Publishing (IFIP) Ltd. with the aim of providing an enterprise-wide publication and integration standard. FEF ontology provides support for modelling financial components and financial entities.

The FP7 FIRST Project (Large Scale Information Extraction and Integration Infrastructure for Supporting Financial Decision Making) has defined an ontology for sentiment analysis in financial domains [9,10]. The ontology identifies Orientation Term (OT), Financial Instrument (FI) and Indicator (I) and their relationships. In addition, the ontology conceptualises specialisations of FI (stocks and stock indexes), economic indicators, and relationships among them. Based on this ontology, the project FIRST has elaborated a set of ontologies for currencies, companies, financial instruments (stocks and stock indexes), funds, financial institutions, insurance companies and banks, available at FIRST project[7].

In its simple form, a FIBO definition would be a single triple. However, FIBO is a complete ontology that enables much more powerful assertions, as will be shown later.

2.2 Linked Opinions and Emotions about stocks

In this section we introduce two vocabularies, Marl and Onyx, that we have defined for providing a uniform vocabulary for expressing sentiments and emotions, respectively, according to linked data principles.

Marl [14] is a standardised data schema designed to annotate and describe subjective opinions expressed on the web or in particular Information Systems. Its aim is to show the benefits of publishing in the open, on the Web, the results of the opinion mining process in a structured form. On the road to achieving

[7] http://first.ijs.si/firstontology/

this, Marl attempts to answer the research question of to what extent opinion information can be formalised in a unified way.

Marl is the result of analysing the properties that characterise opinions expressed on the web or inside various IT systems. The final set of concepts proposed is shown in Fig. 1. It should be noted that opinions in Marl are meant to be linked to an entity. Such entity can be a FIBO Corporation, as described in the previous section. We will make use of this property in Section 2.3.

A detailed description of each particular property and an explanation of their meaning can be found in the vocabulary's specification [8].

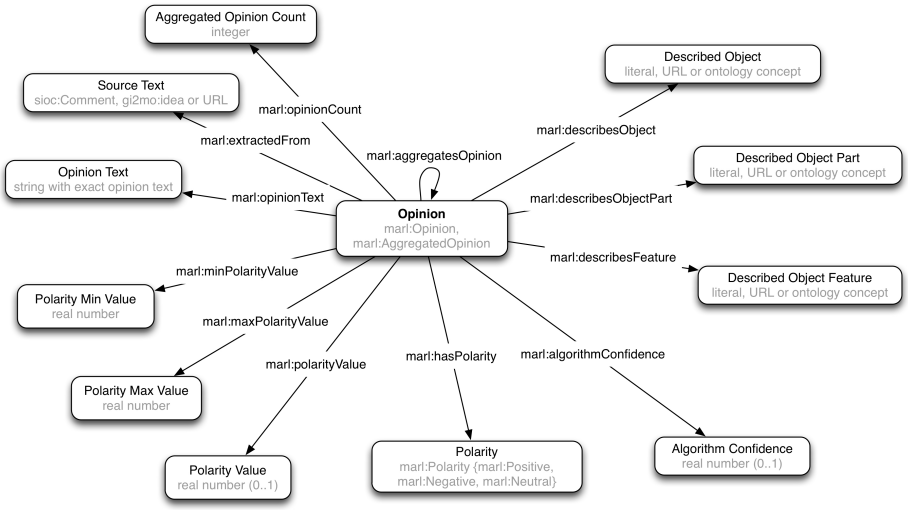

Fig. 1. Marl entities

Onyx [15] is a vocabulary to represent the Emotion Analysis process and its results, as well as annotating lexical resources for Emotion Analysis. It includes all the necessary classes and properties to provide structured and meaningful Emotion Analysis results, and to connect results from different providers and applications.

At its core, the Onyx ontology has three main classes: EmotionAnalysis, EmotionSet and Emotion. In a standard Emotion Analysis, these three classes are related as follows: an EmotionAnalysis is run on a source (generally in the form of text, e.g. a status update), the result is represented as one or more EmotionSet instances that contain one or more Emotion instances.

The specification of the Onyx vocabulary [9] contains an updated description of all its elements, with some usage examples.

[8] `http://www.gsi.dit.upm.es/ontologies/marl`
[9] `http://www.gsi.dit.upm.es/ontologies/onyx`

2.3 Using Linked Data

First of all, let us review a simplified version of the integration of all the elements that we described. To keep it as simple as possible, we will avoid any provenance information (such as who or how analised the twit to extract the opinion) or information about the post itself (author, date, etc.) This simplicity will not prevent us from harnessing the potential of Linked Data.

Listing 1.1. Simple representation using FIBO

```
ex:myOpinion a marl:Opinion;
              marl:hasPolarityValue marl:Positive;
              marl:describesObject ex:GSantander;
              marl:extractedFrom ex:twit1.
ex:twit1 a sioct:MicroblogPost;
         sioc:content "I like testing Grupo Santander".
ex:GSantander a fibo:IncorporatedCompany.
```

In this work, we have gathered thousands of posts from Twitter and stored them in a graph using a more complex version of this schema.

In order to provide a semantic representation of tweets, we have selected TwitLogic [16], which provides a vocabulary for tweets. The basic fields and their relationships are mainly RDF properties and classes taken from well-known sources like FOAF [17] or SIOC [18]. In this work we make use of FIBO to represent the entities of the financial domain. More specifically, we deal with Banks (Incorporated Companies) that have social presence and/or are mentioned by microblogging users. Marl and Onyx have been used for sentiment and emotion annotation, respectively. With this model, we can query all the opinions about a certain entity, statistics such as Positive/Negative ratio, and so on. Listing 1.3 shows an example that gets the count of positive and negative opinions about each entity.

However, the true potential of Linked Data comes into play when we use data from different sources. For instance, if there is another endpoint that contains opinions gathered from Twitter or other social networks, we can query their information seamlessly, provided they use Marl and FIBO as well.

If that example still seems uninteresting, we can also use disparate sources, such as DBpedia. DBpedia contains general information about many entities, which includes several corporations. To be able to query DBpedia, we just need to link our entities to a DBpedia entity. If we take our former example, this modification is as simple as:

Of course, this also involves named entity recognition techniques, which are covered in Section 3.2. Once this step is done, we can issue complex queries that answer questions such as: "What is the general opinion about Banks in Spain?", or "What is the relationship between year of incorporation and the number of opinions in social media?". Note that such queries could use advanced FIBO information, such as current contracts or date of incorporation.

Listing 1.2. Linking FIBO entities to DBpedia

```
ex:GSantander rdfs:seeAlso dbpedia:Santander_Group .
```

Listing 1.3. Query all positive opinions

```
PREFIX sioc: <http://rdfs.org/sioc/ns#>
PREFIX marl: <http://www.gsi.dit.upm.es/ontologies/marl/ns#>
SELECT ?entity
    COUNT(?negative_opinion) AS ?negative_opinions
    COUNT(?positive_opinion) AS ?positive_opinions
WHERE {
{
    ?positive_opinion marl:describesObject ?entity .
    ?positive_opinion marl:hasPolarity marl:Positive .
} UNION {
    ?negative_opinion marl:describesObject ?entity .
    ?negative_opinion marl:hasPolarity marl:Negative .
} } GROUP BY ?entity
```

3 Financial Twitter Tracker Architecture

In this section we describe the architecture of a prototype, called Financial Twitter Tracker, that we have developed for tracking the sentiment evolution of financial entities in Twitter. The core of the system is a semantic pipeline, described below, where tweets are retrieved and analysed. As a result, tweets are semantically annotated as stored in the semantic store Linked Media Framework (LMF) [19]. LMF also provides indexing capabilities based on Solr [20] full text indexing scalable solution. Finally a linked data visualisation framework called Sefarad[10] has been used in order to provide business analysts with a dashboard that assists them in their business decisions, as shown in Fig. 3.

The semantic pipeline for sentiment analysis consists of three tasks. First, the system connects to the Twitter API (Sect. 3.1) and retrieves tweets that match a list of predefined keywords. Then, a semantic analysis (Sect. 3.2 is carried out. Finally the sentiment analysis is done (Sect. 3.3).

3.1 Tweet retrieval

For the purpose of obtaining tweets we developed a wrapper over the services offered by the public Twitter API[11], concretely method search bounded by dates and keywords, which allows the retrieval of each and every tweet published within a particular day and regarding a particular topic. Given the data set of study, several related topics to financial world – such as banking, telecommunication, energy, to name a few – were established. Such data sets have been split according to different languages in order to increase performance and accuracy within the developed "sentiment analysis".

[10] Available at http://github.com/gsi-upm/Sefarad
[11] https://dev.twitter.com/docs/api/1.1/get/search/tweets

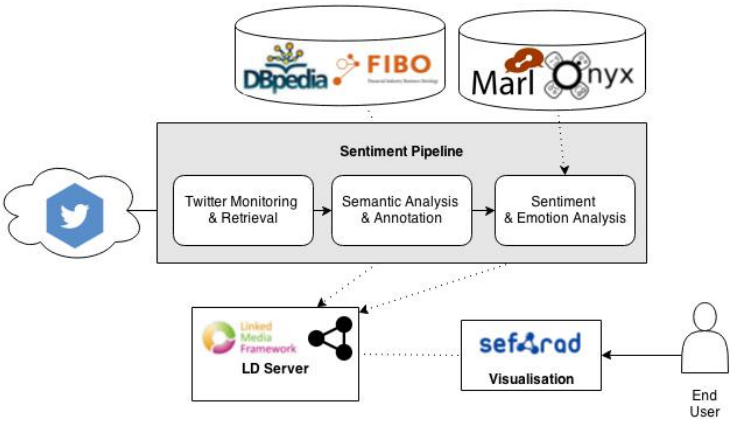

Fig. 2. Financial Twitter Tracker Architecture

3.2 Semantic Analysis and annotation

Data from Twitter is very heterogeneous, as it is used for different purposes (e.g. reviews, factual data, personal comments), covering different categories and subjects. Hence, it was necessary to carry out a categorization prior to the data analysis itself. With such filtering in mind, the Support Vector Machine system (SVM) was developed, taking into account the fact that it supports high dimensional data [21] and their suitability for classifying high volume of information using only support vectors which can be used in any distributed system [22,23,24] offering a great capability of cohesion and adaptation for the MapReduce paradigm. Several studies have proved that SVM provides better results than other techniques of classifications [25]. The system mentioned above has been trained throughout a random sampling of tweets tagged manually using Python with scikit [26] and numpy [27].

As POS-tagging, Treetagger [28] was chosen since it provides support for several languages. After acquiring a financial corpus for tracking a set of financial institutions, this corpus was cleaned, leaving aside irrelevant terms and stop words. Afterwards, collocations were extracted from the most frequent terms generating triplets with a structure domain-context-word (i.e. finance - profits - increasing). Once established these triplets, the following stage was to manually tag them by assigning a quantitative score to determine polarity and synset corresponding to WordNet 3.0 [29][10] basis. These triplets entitle the system to register texts providing scores thanks to the arrangements with WordNet, and leaning on MultiWordNet [30], WN-Affect [31], WN-Domains[32] and SentiWordNet[33]. For this goal, SentiWordNet has been extended in order to reasign scores for the finance domain. The method to enrich the lexicon stands out because its simplicity in terms of configuration, granting the chance of adding new languages easily or extending attached features (affects, domains, scores, etc.)

Another relevant aspect about the lexicon enrichment for its later storage and visualization was the extraction of entities by a NER based on Wikipedia, so that information is compared to the entities published by Wikipedia in order to work out the possible extraction from the text. Periodically the system brings the available information up to date with the new entries published on the online encyclopedia. Finally, that information is lined up with the financial ontology FIBO to provide data in a standardized way in accordance with the semantic web principles such as RDF/OWL, allowing the integration in other technical systems that adapts the given standard. Thanks to FIBO it is possible to provide a clear meaning - without ambiguities - for the financial terms.

3.3 Sentiment and Emotion Analysis

The last stage of the pipeline is in charge of the sentiment and emotion analysis. With a view to quantify the "sentiment" the procedure is to perform the arithmetic mean considering all the registered values recognized in the tweet and using simple rules like inverters (i.e. not). The emotion field can be extracted from the connection between triplets (aligned with WordNet 3.0) and WN-Affect. The outcome stems from the analysis of each tweet which was stored in a MongoDB NoSQL data base, which can handle high volume of information fulfilling the big data requirements of twitter processing.

3.4 Storage and visualisation

After the processing is done, all the triples are stored in an LMF instance, which provides SPARQL and Solr [20] endpoints. We built a generic visualisation framework, Sefarad, that uses these endpoints to display relevant information in any modern browser. This framework is modular and highly customisable. It already contains several plugins that use the power of D3 [12] to display the financial information in several ways. The plugins used, their configuration and location can be configured via an in-browser editor. One of its plugins allows the representation of public sentiment about each entity using Chernoff faces [34].

4 Experimentation

Throughout classification and Sentiment Analysis stages stages of the aforementioned pipeline, we performed experimentation with the obtained data. The classification step has been developed with an SVM trained for the recognition of two groups; finance and non-finance; which states whether the tweets are to continue to the next flow level or, on the contrary, are to be discarded.

Within Machine Learning there are two main discovery methods: supervised and unsupervised learning. In supervised learning, a series of manually tagged data are provided for the system training. On the unsupervised setting, it is the

[12] http://d3js.org/

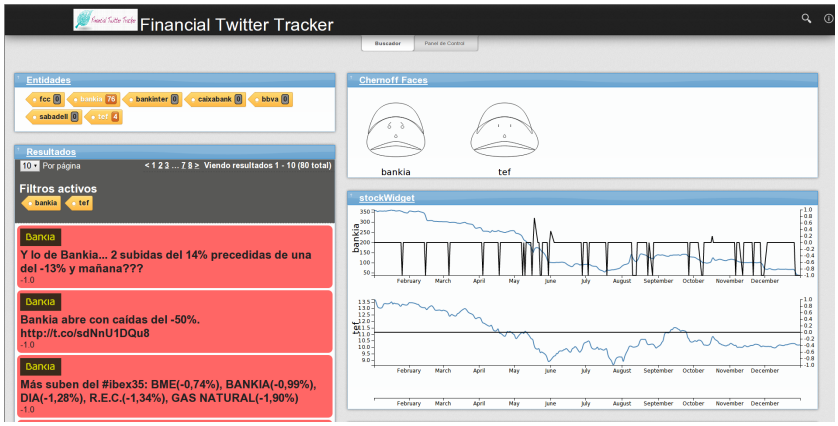

Fig. 3. Financial information displayed using Sefarad

system itself that directly infers patterns from the raw information. The current project uses supervised learning: a random set of tagged tweets has been trained by experts in finances and added to the established groups.

The model has been trained 4 times by modifying the range of information in order to measure and test the system. The first approach makes use of 90% of values for the training and 10% for the assessment, such proportions vary in the second training to 80%-20%, 70%-30% for the third, and 60%-40% for the last one [35]. Each of these cases has been tested five times in order to achieve the harmonic mean of the model accuracy with values chosen randomly for either experiment. The results of these experiments are summarised in Table 1.

Model training-Test	90%-10%	80%-20%	70%-30%	60%-40%
Average precision	0,940	0,9393	0,9369	0,9290
Supported Vectors	886,4548	825,4787	757,501	674,8602

Table 1. Results using different training options

From these results we observe that the bigger the quantity of information used to train the model, the more precise is the outcome, and that value decreases as the volume of data saved for the assessment grows. However, it is remarkable that the more data is used to train the system, the greater is the number of supporting vectors, and, consequently, the classifier loses computational performance.

The Sentiment Analysis phase has been tested against a set of manually annotated corpus. The evaluation was carried out in order to measure the effectiveness. Such accuracy conforms the ability the system owns to satisfy the feature it was developed for [36]. We have classified the results according to the parameters in Table 2. We used these values to define a set of quality metrics as shown in Table 2. The obtained results can be seen in Table 3.

System \ Expert	Identified	Not identified
Retrieved	a	b
Not retrieved	c	d

$$Recall := \frac{a}{a+c} \quad (1)$$

$$Precision := \frac{a}{a+b} \quad (2)$$

$$P_{omissions} := \frac{c}{a+c} \quad (3)$$

$$P_{falsepositive} := \frac{b}{b+d} \quad (4)$$

$$F_1 := 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Table 2. Parameters used in the quality metrics.

Precision	Recall	Probability omissions	Probability false positive	F-Score
84'4%	63'87%	36'12%	77'04%	0.7271

Table 3. Resulting metrics.

5 Conclusions and Future Work

In this article we have presented a vocabulary for modelling sentiments and emotions. This vocabulary can be used to query opinions and emotions about financial institutions and stock values across different web sites and information sources. The main advantage of this approach is that heterogeneous sentiment indexes can be easily integrated and used together with other vocabularies such as FIBO. We have evaluated these vocabularies in a sentiment analysis service based on Twitter for tracking financial institutions.

As a future work, we are working on improving the visualisation and query capabilities of the interface so that non technical users, such as business analysts can take advantage of the possibilities that the Web of Data brings for exploring and consulting, sentiment about financial institutions in large amounts of complex and heterogeneous data.

Another current line of research is the standardisation of these vocabularies for sentiment and emotion. With this aim, we are participating in the Linked Data Models for Emotion and Sentiment Analysis W3C Community Group, which takes as a baseline the vocabularies Marl and Onyx.

6 Acknowledgement

This research has been partially funded by the Spanish Ministry of Industry, Tourism and Trade through the project Financial Twitter Tracker (TSI-090100-2011-114) and the EUROSENTIMENT FP7 Project (Grant Agreement no: 296277)

References

1. Chatfield, A.T., Brajawidagda, U.: Twitter early tsunami warning system: A case study in indonesia's natural disaster management. In: System Sciences (HICSS), 2013 46th Hawaii International Conference on. (Jan 2013) 2050–2060
2. Yardeni, E.: Stock market indicators: Fundamental, sentiment & technical. Technical report, Yardeni Research (2014) Available at `http://www.yardeni.com/pub/peacockbullbear.pdf`.
3. Vu, T.T., Chang, S., Ha, Q.T., Collier, N.: An experiment in integrating sentiment features for tech stock prediction in twitter. In: 24th International Conference on Computational Linguistics. (2012) 23
4. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science **2**(1) (2011) 1–8
5. O'Riain, S., Curry, E., Harth, A.: XBRL and open data for global financial ecosystems: a linked data approach. International Journal of Accounting Information Systems **13**(2) (June 2012) 141–162
6. O'Riain, S., Harth, A., Curry, E.: Linked Data Driven Information Systems as an Enabler for Integrating Financial Data. In: Information Systems for Global Financial Markets: Emerging Developments and Effects. IGI Global (2012)
7. Krieger, H., Declerck, T., Nedunchezhian, A.: MFO - the federated financial ontology for the monnet project. In: Proceedings of the 4th International Conference on Knowledge Engineering and Ontology Development, Barcelona, Spain (2012)
8. O'Riain, S., Coughlan, B., Buitelaar, P., Declerk, T., Krieger, U., Marie-Thomas, S.: Cross-lingual querying and comparison of linked financial and business data. In: Proceedings of 10th Extended Semantic Web Conference (ESWC), Montpellier, France (2013)
9. Grcar, M., Häusser, T., Ressel, D.: D3.1 semantic resources and data acquisition. Technical report, First project (2011)
10. Klein, A., Altuntas, O., Häusser, T., Kessler, W.: Extracting investor sentiment from weblog texts: A knowledge based approach. In: Proc. of the 2011 IEEE Conference on Commerce and Enterprise Computing. (2011) 1–9
11. Goto, M., Hu, B., Naseer, A., Vandenbusshe, P.I.: Linked data for financial reporting. In: 4th International Worshop on Consuming Linked Data (COLD2013), CEUR Workshop proceedings (2013)
12. Council, E.: FIBO. Financial Industry Business Ontology. Available at http://www.edmcouncil.org/financialbusiness (June 2013)
13. IFIP: FEF. financial exchange framework ontology. Available at http://www.financial-format.com/index.html (June 2003)
14. Westerski, A., Iglesias, C.A., Tapia, F.: Linked Opinions: Describing Sentiments on the Structured Web of Data. In: Proceedings of the 4th International Workshop Social Data on the Web. (2011)
15. Sánchez-Rada, J.F., Iglesias, C.A.: Onyx: Describing Emotions on the Web of Data. In: Proceedings of the 1st International Workshop on Emotion and Sentiment in Social and Expressive Media: approaches and perspectives from AI, Torino, Italy, AI*IA, Italian Association for Artificial Intelligence (December 2013)
16. Shinavier, J.: Real-time# semanticweb in $\leq$ 140 chars. In: Proceedings of the Third Workshop on Linked Data on the Web (LDOW2010) at WWW2010. (2010)
17. Golbeck, J., Rothstein, M.: Linking social networks on the web with foaf: A semantic web case study. In: AAAI. Volume 8. (2008) 1138–1143

18. Breslin, J.G., Decker, S., Harth, A., Bojars, U.: Sioc: an approach to connect web-based communities. International Journal of Web Based Communities **2**(2) (2006) 133–142
19. Kurz, T., Schaffert, S., Burger, T.: Lmf: A framework for linked media. In: Multimedia on the Web (MMWeb), 2011 Workshop on, IEEE (2011) 16–20
20. Smiley, D., Pugh, D.E.: Apache Solr 3 Enterprise Search Serve. Packt Publishing (2011)
21. Dilrukshi, I., De Zoysa, K., Caldera, A.: Twitter news classification using svm. In: Computer Science & Education (ICCSE), 2013 8th International Conference on, IEEE (2013) 287–291
22. Jakkula, V.: Tutorial on support vector machine (svm). School of EECS, Washington State University (2006)
23. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: LREC. (2010)
24. Balahur, A., Steinberger, R., Goot, E.v.d., Pouliquen, B., Kabadjov, M.: Opinion mining on newspaper quotations. In: Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on. Volume 3., IET (2009) 523–526
25. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics (2002) 79–86
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
27. Oliphant, T.E.: Guide to NumPy, Provo, UT. (March 2006)
28. Schmid, H.: Probabilistic part-of-speech tagging using decision trees. In: Proceedings of international conference on new methods in language processing. Volume 12., Manchester, UK (1994) 44–49
29. Fellbaum, C.: WordNet. Wiley Online Library (1999)
30. Pianta, E., Bentivogli, L., Girardi, C.: Multiwordnet. developing an aligned multilingual database. In: Proc. 1st International Conference on Global WordNet. (2002)
31. Strapparava, C., Valitutti, A.: Wordnet affect: an affective extension of wordnet. In: LREC. Volume 4. (2004) 1083–1086
32. Magnini, B., Strapparava, C., Pezzulo, G., Gliozzo, A.: The role of domain information in word sense disambiguation. Natural Language Engineering **8**(4) (2002) 359–373
33. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: LREC. Volume 10. (2010) 2200–2204
34. Chernoff, H.: The use of faces to represent points in k-dimensional space graphically. Journal of the American Statistical Association **68**(342) (1973) 361–368
35. Kohavi, R., et al.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI. Volume 14. (1995) 1137–1145
36. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval. Volume 1. Cambridge university press Cambridge (2008)